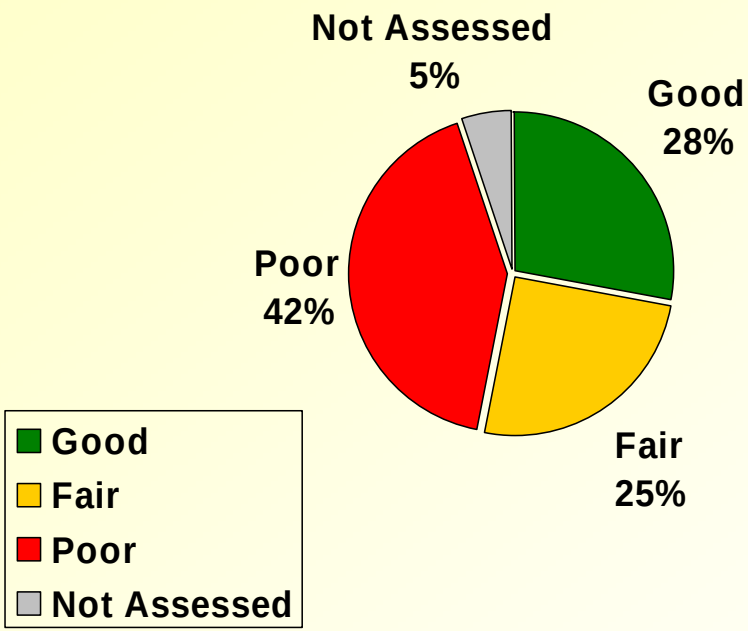
Comparability and Integration of State-, Regional-, and National-Scale Probability-Based Surveys

Anthony (Tony) R. Olsen
Aquatic Monitoring and Assessment Branch
Western Ecology Division, NHEERL
Office of Research and Development
U.S. EPA, Corvallis, OR
Email: olsen.tony@epa.gov

April 25, 2006

Hypothetical National Estimate



The National Lake Assessment found that 28% of lakes in United States are in good condition.

Getting National Assessments Based on State Assessments

$$\hat{L}_{us} = \hat{L}_{al} + \hat{L}_{ar} + \ldots + \hat{L}_{wy}$$

$$\hat{P}_{us} = \left(\frac{LF_{al}}{LF_{us}}\right) * \hat{P}_{al} + \left(\frac{LF_{ar}}{LF_{us}}\right) * \hat{P}_{ar} + \ldots + \left(\frac{LF_{wy}}{LF_{us}}\right) * \hat{P}_{wy}$$

$$Var(\hat{P}_{us}) = \left(\frac{LF_{al}}{LF_{us}}\right)^2 * Var(\hat{P}_{al}) + \left(\frac{LF_{ar}}{LF_{us}}\right)^2 * Var(\hat{P}_{ar}) + \ldots + \left(\frac{LF_{wy}}{LF_{us}}\right)^2 * Var(\hat{P}_{wy})$$

where

$\hat{L}_{xx}$ = estimated number of lakes in "xx" state or U.S.

P_{xx} = estimated proportion of lakes in good condition for "xx" state or U.S.

LF_{xx} = Number of lakes in NHD (sample frame) for "xx" state or U.S.

$Var(\hat{P}_{xx})$ = variance of estimated proportion of lakes in good condition for "xx" state or U.S.

What Can Cause Problems?

- Objective: Focus on long-term goal of cost-effective, defensible assessments of lakes at state, tribal, regional, and national scales.
- Target Population: Understanding state, tribal, regional, and national lake definitions and impact on design.
- Sample Frame: Importance of sample frame development and integrating them across organizations.
- Survey Design: Using probability-based survey designs that meet needs at all scales.
- Response Design: Incorporating indicators relevant to each organization.
- Site Evaluation: Adopting common process useful at all scales.
- Population Estimation: Learning how to make lake assessments at all scales compatible.

Target Population Definitions

- Desirable: A single definition that all agree upon: A lake/pond/reservoir is…"set of criteria"
- Necessary:
  - Explicit set of agreed-upon criteria that identifies a lake
  - Documentation by each organization on which criteria used in their definition
  - Lake evaluation information for each lake giving status of lake for each criteria

Possible Definition Criteria

- **Desirable example:** A lake/pond/reservoir is a water body that:
  - is a permanent water body
  - has surface area greater than or equal to 1 hectare
  - is not a saline water body
  - is not used for aquaculture, disposal tailings, sewage treatment, evaporation, or other unspecified disposal use
  - is greater than 1 meter deep
  - has at least 1,000 square meters surface area of open water
- **Criteria example:**
  - permanent versus seasonal water body
  - minimum surface area
  - saline versus non-saline
  - special use water body types
  - minimum depth
  - minimum open water surface area

Why is This Important?

- To compare or combine surveys by different organizations, we must be able to create a lake definition that is common to all.
- National criteria are “lowest common criteria” unless option available to survey lakes not sampled by some states or tribal nations

“Lurking” Definition Issue

- Is a lake . . .
  - assessed as a single unit (discrete entity)?
  - assessed separately in different portions of the lake (assessment units)?
  - assessed continuously (areal entity)?
- Can assessments based on discrete entities, assessment units, and areal entities be combined?

Sample Frame Issues

- Sample frame is best GIS representation of lakes that meet the target population definition
- Each organization creates a sample frame that meets their requirements
- When integrating surveys, it is critical that a single combined sample frame be constructed
- Must know if all sample frames include lakes of interest
- Sample frames must have common set of attributes necessary when combined nationally

Survey Design Issues

- Every state and tribal nation either must use a probability-based survey design OR sample every lake within their jurisdiction.
- If national estimate is based on lakes sampled in one year (*e.g.*, 2007),
  - then survey design in each state must cover entire state in 2007
  - *i.e.*, a state using a rotating basin design would not have necessary design for national estimate
- If national estimate is based on data from past 5 years,
  - then rotating basin designs would provide necessary data
- National survey requires estimates by aggregated ecoregions
  - sample size required would need to be assured by collection of state sample sizes

Response Design Issues

- Does definition of “good” condition mean the same thing for all states?
- Do states measure a core set of indicators that are required for national estimates?
- Are indices, such as IBI, calculated the same?
- Do states use same field and lab protocols for an indicator?

Site Evaluation Issues

- **Because sample frames based on NHD are not perfect lists for all lakes in target population, some lakes selected will not be sampled**
  - Are states keeping track of why a lake was not sampled?
  - Are states using same, or comparable, information to make a decision not to sample?
- **Some lakes require land owner permission to sample**
  - Are states using similar procedures to gain access?
  - Are the procedures designed to minimize access denials?
- **Some lakes are physically unsafe or difficult to access**
  - Are states using similar procedures/effort to get access?
  - Access is more difficult in some states (*e.g.,* Alaska) than others. How does this impact national estimates?

Population Estimation Issues

- Integrating data from state surveys is an information management task
  - What information is required to do national estimates?
  - When is the information required?
- How are national estimates produced
  - Aggregation of state-level estimates?
  - Estimates based on state-level data?

Fundamental Design Options

- National-level survey design with provision for 50 state-level intensifications
- 50 state-level survey designs that are coordinated to ensure national estimates
- Near-term national-level survey design necessary
- Long-term objective is integrated state-level and national-level survey design

http://www.epa.gov/nheerl/arm